The FHWA Travel Model Improvement Program Workshop over the Web

The Travel Model Development Series: Part I – Travel Model Estimation

presented by
Thomas Rossi
Yasasvi Popuri
Cambridge Systematics, Inc.

March 12, 2009

1

Homework
From Session 4

2

Webinar Outline

- Session 1: Introduction – October 16, 2008
- Session 2: Data Set Preparation – November 6, 2008
- Session 3: Estimation of Non-Logit Models – December 11, 2008
- Session 4: Estimation of Logit Models – February 10, 2009

3

Webinar Outline – Note Revisions! (continued)

- Session 5: Disaggregate and Aggregate Validation Procedures – March 12, 2009
- Session 6: Advanced Topics in Discrete Choice Models – April 14, 2009*
- Session 7: Highway and Transit Assignment Processes – May 7, 2009

4

Webinar Outline – Note Revisions! (continued)

- Session 8: Evaluation of Model Validation Results – June 9, 2009
- NEW SESSION – Session 9: Real Life Experiences in Model Development, Webinar Wrap-Up – July 16, 2009

5

Note on Session 6

Session 6: Advanced Topics in Discrete Choice Models – April 14, 2009

- This is an optional session, requested by reviewers of the original webinar outline
- More detail, more math on logit models
- No homework
- Therefore, Session 5 homework will be reviewed at the beginning of *Session 7*

6

The Model Validation Process

- We will be discussing the overall model validation process in Session 8, but…
- One of the key concepts in model validation is that each component of a model must be validated individually
- This session deals with validating the various types of models we have seen so far in the webinar

7

The New FHWA Model Validation Manual Is Coming!

- Later this year!
- Rejected titles for the new manual:
  - "Son of Model Validation and Reasonableness Checking Manual"
  - "Validation Redux!"
  - "Validation II – The Sequel!"
  - "Validation Wars Episode 5 – The Modeler Strikes Back"

8

Validation Includes a Lot of Things

- Checks of input data
- Reasonableness/logic checks
- Comparison of model results to independent data sources
- Sensitivity checks

9

Aggregate vs. Disaggregate Validation

- Disaggregate validation
  - Explores how well model fits observed data at the household or individual level
  - Involves defining subgroups of observations
  - Compares model results with observed data to reveal systematic biases
  - Plays more of a role in the model estimation phase
- Aggregate validation
  - Provides a general overview of model performance through regional travel characteristics
  - Applies model at the regional, district, and zonal level

10

Aggregate vs. Disaggregate Validation

- Aggregate models require aggregate validation
- Disaggregate models require both aggregate and disaggregate validation

11

Aggregate Validation

- Assumes checks of model estimation have been done at time of estimation
- Generally involves applying models to perform reasonableness checks
- Comparison of model results to independent data sources
  - Remember, comparison is not always “matching”
- Looks at overall results and results by market segment

12

Cross-Classification Model – Trip Productions
Review

		Independent Variable #1				
		Value 1	Value 2	…	Value n	Total
Independent Variable #2	Value 1	Dep var value	Dep var value		Dep var value	
	Value 2	Dep var value	Dep var value		Dep var value	
	…					
	Value n	Dep var value	Dep var value		Dep var value	
	Total					

13

Checks of Trip Production Rates

- Comparisons to other sources
  - Other models
  - NHTS
  - NCHRP Report 365 and updates
- Marginal totals

14

Aggregate Trip Rates from Application

- Trips per household
- Trips by purpose
- Application to other than year of estimation data

15

Calibration of Trip Rates

- Where are the discrepancies?
- Are the discrepancies really wrong?
- Checking the input data

16

Regression Model – Attractions Review

$Y = B_0 + B_1 X_1 + B_2 X_2 + \ldots + B_n X_n$

where:

Y = Dependent variable

B_i = Estimated coefficients

X_i = Independent variables

The maximum likelihood estimators for coefficients are based on method of least squares

17

Aggregate Trip Rates from Application

- Trips per employee (by type)
- Trips by purpose
- Comparison to trip productions
- Application to other than year of estimation data

18

Calibration of Parameters

- Where are the discrepancies?
- Are the discrepancies really wrong?
- Checking the input data

19

Gravity Model - Trip Distribution Review

$$T_{ij} = \frac{P_i A_j F(t)_{ij} K_{ij}}{\sum_j P_i A_j F(t)_{ij} K_{ij}}$$

where:

- T_{ij} = number of trips produced in zone i and attracted to zone j
- P_i = trips produced in zone i
- A_j = trips attracted to zone j
- $F(t)_{ij}$ = friction factor from i to j (based on impedance t)
- K_{ij} = K factor from i to j
- i = origin zone
- j = destination zone

20

Trip Length Frequency Distribution
The First (but not only) Check

- Use skims for both observed and model results
- Check averages and fit
- Check by market segment
- Application to other than year of estimation data

21

Geographic Distribution

- Compare to expanded survey data
- District-to-district trips
- Intrazonal trips
- Application to other than year of estimation data

22

Gravity Model Calibration

- Trip length differences
  - Adjust friction factors or parameters from function
- Geographic differences
  - When are K-factors OK?
- As always, check input data

23

Logit Models
Review

Probability function:

$$P(1) = \frac{\exp(v_1)}{\exp(v_1) + \exp(v_2) + \cdots + \exp(v_n)}$$

Used for:

- Mode choice
- Vehicle availability
- Destination choice
- And others…

24

Disaggregate Validation

- Two ways of doing this
  - Apply model to a data set independent of the estimation data set
  - Apply model to the estimation data set, report results by market segment

25

Disaggregate Validation (continued)

- Application to original data set, market segments
  - Household characteristics such as household size, income level, auto ownership, etc.
  - Traveler characteristics such as age, gender, driver's license status, and employment status
  - Zonal characteristics such as geographical location, area type, etc.
  - Trip characteristics such as trip distance, time, and cost

26

Example of Disaggregate Validation

Choice	0	1	2	3+	Total
Non Motorized					
Number Chosen	47.4	104.5	270.2	158.5	580.6
Standard Deviation Chosen	7.5	14.1	26.5	19.0	36.3
	*A	V	A	A	A
Number Predicted	32.8	117.4	264.5	150.7	565.2
Auto Passenger					
Number Chosen	40.5	277.3	537.8	351.8	1,207.5
Standard Deviation Chosen	7.2	18.1	32.3	27.3	46.6
	V	***A	V	*V	A
Number Predicted	47.1	197.7	549.8	386.1	1,180.7
Drive Alone					
Number Chosen	0.0	1,265.9	4,225.5	3,233.4	8,724.8
Standard Deviation Chosen	0.0	25.4	44.8	35.7	62.7
		**V	A	A	A
Number Predicted	0.0	1,317.44	4,204.4	3,201.1	8,723.0
⋮	⋮	⋮	⋮	⋮	⋮
Total					
Number Chosen	119.3	1,770.7	5,326.2	3,928.7	11,144.9
Number Predicted	119.3	1,770.7	5,326.2	3,928.8	11,144.9

27

Aggregate Logit Model Checks
Example: Mode Choice Model

- Mode shares by purpose and market segment
- Comparison of transit trips to results from on-board survey
  - Origin-destination
  - Trip purpose
  - Rider demographics

28

Aggregate Logit Model Checks
Example: Mode Choice Model (cont'd)

- Transit assignment checks
  - Line/station boardings
  - Corridor volumes
  - Screenlines
  - Transfers

29

Mode Choice Calibration

- More than just regionwide validation – "adjusting constants"
- Segmentation variables – revising, adding, deleting
- Adjusting network parameter and settings
- Often "points back" to issues with earlier model steps

30

Vehicle Availabilty Model Validation Example

Variable	County							
	Yasasvi	Sarah	Gary	Tony	Chris	Sam	Lori	Region
% 0 Vehicles								
Observed	5%	13%	12%	7%	12%	7%	37%	17%
Model	3%	8%	8%	3%	5%	4%	24%	11%
% 1 Vehicles								
Observed	28%	35%	35%	31%	34%	33%	42%	35%
Model	27%	38%	38%	30%	34%	33%	49%	37%
% 2 Vehicles								
Observed	44%	37%	38%	44%	38%	43%	18%	34%
Model	47%	39%	39%	45%	43%	43%	21%	36%
% 3 Vehicles								
Observed	22%	14%	15%	18%	16%	17%	3%	13%
Model	22%	15%	15%	21%	18%	20%	5%	15%
Average Vehicles								
Observed	1.86	1.55	1.57	1.78	1.58	1.74	0.88	1.44
Model	1.95	1.65	1.65	1.90	1.79	1.82	1.11	1.59

31

Vehicle Availability Model Calibration Example

- Check 0-vehicle households – data set, observed data
- Check county level validation
  - Why is VA overestimated in Chris and Lori Counties?

32

Sensitivity Checks

- Ensure that sensitivity of model outputs to changes in inputs is reasonable
- Apply model with known changes in inputs
  - Socioeconomic characteristics (growth)
  - For mode choice, time and cost

33

Homework
Session 5

34